Prediction of the Composition of Coal Tars from the Pyrolysis Mass Spectra of the Parent Coals Using Canonical Correlation Techniques

Tanmoy Chakravarty, Henk L.C. Meuzelaar, Patrick R. Jones

Biomaterials Profiling Center, University of Utah,
Salt Lake City, UT 84102

M. Rashid Khan

Morgantown Energy Technology Center, Morgantown, W. Virginia

INTRODUCTION

In this study, we show that the mass spectrometric composition of coal tars can be predicted from the pyrolysis mass spectra of their parent coals. Curie-point low voltage electron ionization mass spectrometry (CuPy-EIMS) was performed on nineteen coals and on their respective pyrolysis liquids prepared by means of a fixed bed reactor method described elsewhere (1). Using factor and discriminant analysis techniques, the spectra can be classified and underlying structural variables responsible for the above classification identified. Furthermore, compositional similarities and dissimilarities between the solid samples and their liquids can be brought out using canonical correlation methods (2). Table I lists the samples, their PSOC numbers, geological origin and the rank information. Of the nineteen coal samples, two are of subbituminous rank, three of high volatile B and C bituminous rank and the remainder of high volatile A bituminous rank. Twelve coals are from the Eastern/Appalachian coal province, six are from the Interior province and one is a Western coal from the Northern Great Plains province.

EXPERIMENTAL

Details of sample preparation and CuPy-EIMS analysis procedures have been described elsewhere (3). Experimental conditions were as follows: Curie-point temperature 610°C; heating rate approx. 100 K/s; total heating time 10s; electron energy 12eV; mass range scanned 20-260 amu; scanning rate 1000 amu s^{-1}; total scan time approximately 10s. Each sample was analyzed in triplicate and the spectra recorded and stored by computer (IBM 9000).

Data Processing

Mass variables with m/z values higher than 90 were selected for multivariate analysis. The reason for this is that the low mass range variables are not suitable for correlating the coals with their tars because of the loss of low molecular weight volatile products in the fixed bed reactor. The remaining data sets consisting of 171 variables were preprocessed, normalized and factor and discriminant analysis performed (4,5). Four discriminant functions were found to be significant, accounting for 55% of the variance in the coal data and 53% variance in the tar data, respectively.

Using four discriminant functions from both the coal and the tar data sets, canonical correlation analysis was done. Two canonical variates were obtained with correlation coefficients >0.9. The two canonical variates for the coal data set accounted for 35% of the total variance; whereas 30% of the total variance was accounted for by the two canonical variates for the tar data set.

Using a "jackknifing" procedure the topology of the canonical variate space was checked. The scores of the "unknown" sample were projected in the canonical variate space and the pyrolysis mass spectra of these unknown samples were then

predicted by calculating a distance weighted average from the spectra of the two nearest neighbors.

Methods and Procedures

A brief description of the methodology and mathematical rationalization is given below. Factor analysis (6) is an efficient way of reducing a data set (the data is generally preprocessed to form a correlation matrix). The first factor describes the maximum correlated variance in the data set, the second (orthogonal to the first) the maximum of the residual correlated variance, etc.:

$$F_j = a_{1j}z_1 + a_{2j}z_2 + \ldots + a_{mj}z_m \qquad 1)$$

where F_j is factor j, with loading a_{ij}, which describe the contribution of the variable z_i to the factor.

The scores, i.e., the contributions of the spectra to the factors, are obtained by substituting the intensities of the mass variables in the spectra for z_i. The relationship between the data matrix D (size s x m where s is the number of spectra and m is the number of mass variables) and the calculated factor matrix is:

$$D = S \times F \qquad 2)$$

where S is the score matrix (size s x n, n = number of factors) and F (size n x m) contains the factors. The scores in the principal component analysis are calculated in the following way:

$$S = D \times (E \times /\backslash^{-1/2}) \qquad 3)$$

where S contains the standardized scores (mean is zero, standard deviation is 1), and D the standardized data matrix. E is the orthonormal eigenvector for matrix Z = 1/s D^TD, with the eigenvalues given in /\.

The standardized data matrix, D, can be reconstructed from the scores using this relation:

$$\begin{aligned} D &= S \times (E \times /\backslash^{-1/2})^{-1} \\ &= S \times (E^{-1} \times /\backslash^{1/2}) \\ &\square\ S \times (E^T \times /\backslash^{1/2}) \\ &= \text{standardized scores} * \text{factor loadings} \end{aligned} \qquad 4)$$

The score of an unknown sample, defined in terms of a standardized data matrix, can be projected using Equation 3. Similarly, the projected unknown score, based on a training data set, can be used to predict a standardized data matrix (in this case of the unknown sample) using Equation 4. The factor loadings, before being used in Equation 4, have to be transformed to the canonical variate space by using the canonical variate matrix.

RESULTS AND DISCUSSION

The results will be discussed under two separate subheadings; canonical correlation results and prediction results.

Canonical Correlation Results

Canonical variate scores (CVI vs. CVII) for the two data sets are plotted in Figure 1. The result shows that the data is primarily two-dimensional. The clustering of the samples suggests that the Interior hvb coals (samples #15, 16 and

17, Table I) are more alike in composition than the Eastern hvb coals. The subbit C Upper Block Indiana coal (sample #7) is different from the subbit A coal of the Northern Great Plains (sample #18). The Lower Kittaning coal from West Virginia (hvAb, sample #9) and the Lower Banner Virginia coal (also hvAb, sample #6) do not fall close to the cluster with other hvAb coals. Although the Lower Banner coal has been found to be different from other hvAb coals petrographically, similar data is unavailable for the Lower Kittaning coal.

The clustering pattern of the corresponding tar data set is almost the same as that for the coal data; the scores of the coal and the tar are very close in this two-dimensional canonical variate space. The closer the score of the tar sample to that of the coal, the more alike their spectral patterns are. The average relative deviation between the coal and tar scores is 10% in the direction of CVI and 25% in the direction of CVII (this is related to the uncertainty in the predicted composition). Based on this canonical score plot, it is possible to predict the composition of the coal tars starting from the Py-MS patterns of the coals with the help of Equation 4.

As mentioned before, the prediction is based on the location of the projected score of the unknown sample in the canonical variate space of the coal data set. Two samples were selected for jackknifing tests - one from each marked cluster representing coals of different rank and origin (samples #16 and 2, respectively). Figures 2 and 3 show the score plots for the first two canonical variates with the score of the particular unknown sample projected. The topology of the CVI/CVII space is preserved in both cases (Figures 2 and 3); the projected score of the particular sample falls almost in the same location as if it were a part of the data set.

Prediction Results

Because the projected score of the jackknifed sample lies in the same space as the original data set, we have used a simple technique to predict the mass spectra of the selected samples. For example, the spectrum of sample #16 is derived from the distance weighted averaging of those of samples 15 and 17 (Figure 2). Similarly, the spectrum of sample #2 is derived by weighted averaging of those of samples 4 and 14 (Figure 3).

Since some coal and tar components are not strongly represented in this space the distance weighted averaging method makes the assumption that clustering behavior observed in canonical correlation space is representative of overall clustering tendencies. This assumption can and should be verified by inspection of clustering behavior in multidimensional factor space.

Figure 4a, b, c shows the spectra of the IL #6 coal, that of the corresponding tar, and the predicted tar. Most of the components present in the coal spectrum, Figure 4a, are also seen in the tar spectrum, Figure 4b. For example, the mass peaks at m/z 94, 108, 122 ("phenols"), 110, 124, 138 ("dihydroxybenzenes"); 142, 156, 170, 184, 198 ("naphthalenes"); 168, 182, 196, 210, 224 ("acenaphthenes") and 178, 192, 206 ("phenanthrenes/anthracenes") are all found in both the coal and the tar spectra. The differences between these two spectra should also be noted here. The higher molecular weight masses all have stronger signals in the tar spectrum, and the patterns of the components such as naphthalenes (m/z 142, 156, 170, 184), acenaphthenes (m/z 168, 182, 196) and anthracenes (m/z 178, 192, 206) are different in the tar spectrum. The naphthalene series reaches its maximum at at m/z 156 in the coal spectrum, and at m/z 170 in the tar spectrum; similar shifts in the relative intensity patterns are observed for other components mentioned above. The predicted spectrum, Figure 4c, shows that it preserves the spectral subpatterns of the individual components. Also note that the ratio of the peak heights between

and within separate components in the original tar spectrum (Figure 4b) and the predicted spectrum (Figure 4c) is alike.

Figure 5a, b, c shows the spectra of a typical Eastern hvAb coal from Pennsylvania, of the corresponding tar, and of the predicted tar. In this case the spectral subpatterns of the important components are markedly different between the coal and the tar. Nevertheless, when comparing Figures 5b and 5c the predicted and measured tar patterns show a very high degree of similarity and are distinctly different from the corresponding patterns in Figure 4.

CONCLUSIONS

The data presented here demonstrate that pyrolysis mass spectrometry in combination with canonical correlation analysis enables modeling and prediction of complex coal conversion processes such as the fixed bed liquefaction method. To the best of our knowledge this represents the first time that the chemical composition of coal-derived liquids has been predicted directly from feed coal characterization data. Moreover, the method is completely general and can be applied to all coal characterization data (whether obtained by conventional or by advanced spectroscopic techniques) and coal conversion processes in which the end products have been carefully characterized.

A shortcoming of the present study is the lack of a sufficiently large number of feed coal/coal tar pairs to allow adequate modeling and prediction of coal conversion behavior for more than two coal clusters. Also, a somewhat simplistic method was used to calculate predicted tar spectra based on a distance weighted average of the two nearest neighbors in a two-dimensional canonical correlation space.

REFERENCES

1. Khan, M.R., Proc. of Int. Symp. on Coal Science, Coal Science and Technology Series, J.A. Moulijn (ed.), Elsevier, 1987, p. 647.

2. Metcalf, G.S., Windig, W., Hill, G.R., Meuzelaar, H.L.C., Int. J. Coal Geology, 1987, 7, 245-268.

3. Chakravarty, T., Meuzelaar, H.L.C., Report to U.S. Dept. of Energy, July 1987, Contract #DEAP21-87MC05073.

4. Windig, W., Meuzelaar, H.L.C., Anal. Chem., 1984, 56, 2297-2303.

5. Chakravarty, T., Windig, W., Taghizadeh, K., Meuzelaar, H.L.C., Shadle, L.J., Energy & Fuels, 1988 (in press).

6. Malinowski, E.R., Factor Analysis in Chemistry, Academic Press, New York, 1980.

ACKNOWLEDGEMENTS

The authors wish to acknowledge Ms. Barbara Hoesterey for her assistance in sample preparation and analyses. Helpful discussions with Dr. Willem Windig is also appreciated. This work was supported by the Department of Energy (Contract #DEAP21-87MC05073). This work was sponsored in part by the Advanced Combustion Engineering Research Center. Funds for this Center are received from the National Science Foundation, the State of Utah, 21 industrial participants, and the U.S. Department of Energy. Additional funding was also provided by the Consortium for Fossil Fuel Liquefaction Science.

TABLE I
GEOGRAPHICAL ORIGIN AND RANK INFORMATION OF NINETEEN COAL SAMPLES

SERIAL #	PSOC #	GEOLOGICAL INFORMATION	RANK INFORMATION
1	-	Eastern/Appalachian, OH #6	hvAb
2	1481	ibid, Upper Clarion	hvAb
3	-	ibid, Wellmore #8	hvAb
4	375	ibid, Hazard #9	hvAb
5	267	ibid, Clintwood	hvAb
6	1472	ibid, Lower Banner	hvAb
7	-	ibid, Arkwright	hvAb
8	1469	ibid, Mary Lee	hvAb
9	123	ibid, Lower Kittaning	hvAb
10	1471	ibid, Pee Wee	hvAb
11	306	ibid, OH #12	hvAb
12	296	ibid, OH #5	hvAb
13	1475	ibid, Elkhorn #3	hvAb
14	275	ibid, OH #6A	hvAb
15	1492	Interior/Eastern, IL#5	hvBb
16	-	ibid, IL#6	hvCb
17	1323	ibid, IL#6	hvCb
18	181	ibid, Upper Block	sub. A
19	1520	N. Great Plains/Fort Union, Wyodak	sub. C

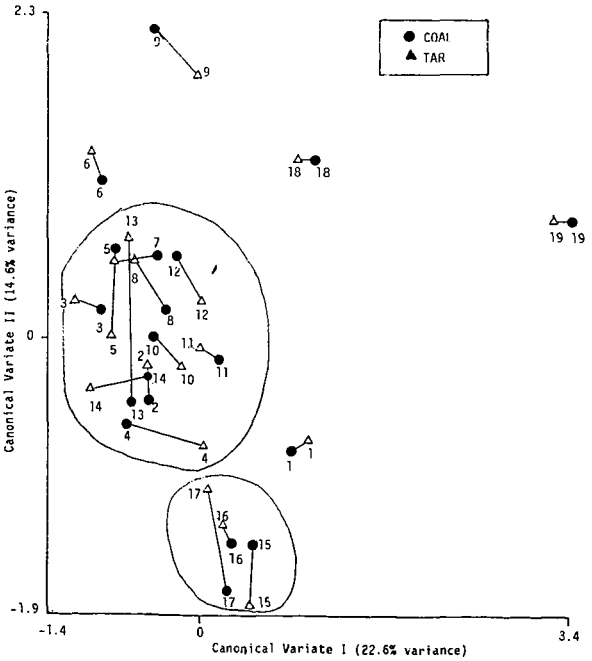

Figure 1. Score plot in the CVI/CVII space for both the coal and the tar samples. Note that only the means of the three scores for each category are plotted.

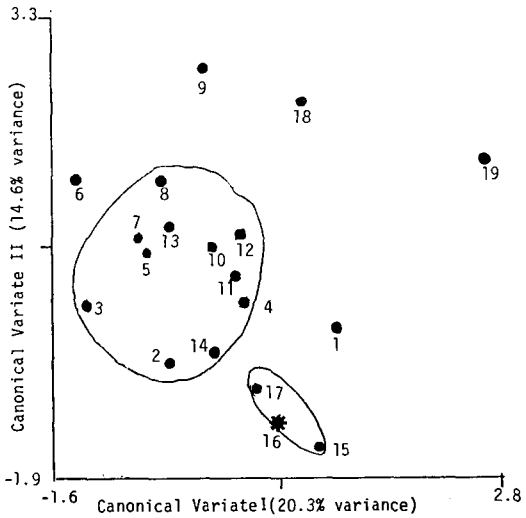

Figure 2. Effect of removing sample #16 from the coal data set on the score plot in the CVI/CVII space. Note the projected score of sample #16 (marked by *).

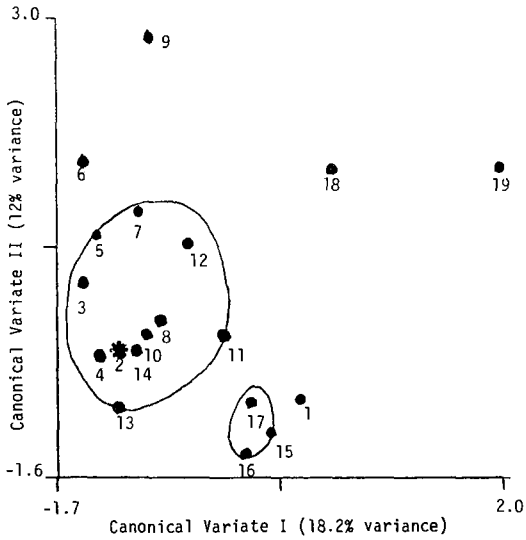

Figure 3. Effect of removing sample #2 from the coal data set on the score plot in the CVI/CVII space. Note the projected score of sample #2 (marked by *).

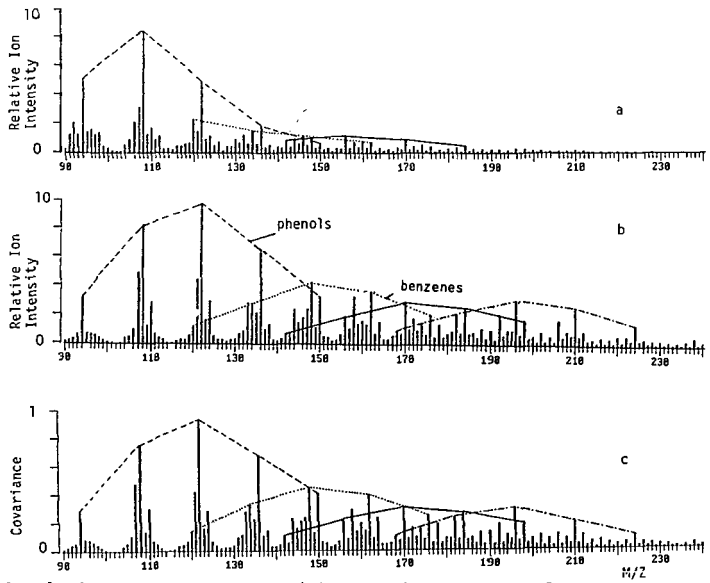

Figure 4. Pyrolysis mass spectrum of: (a) Illinois #6 hvCb coal; (b) direct probe mass spectrum of the corresponding tar, and (c) mathematically predicted spectrum.

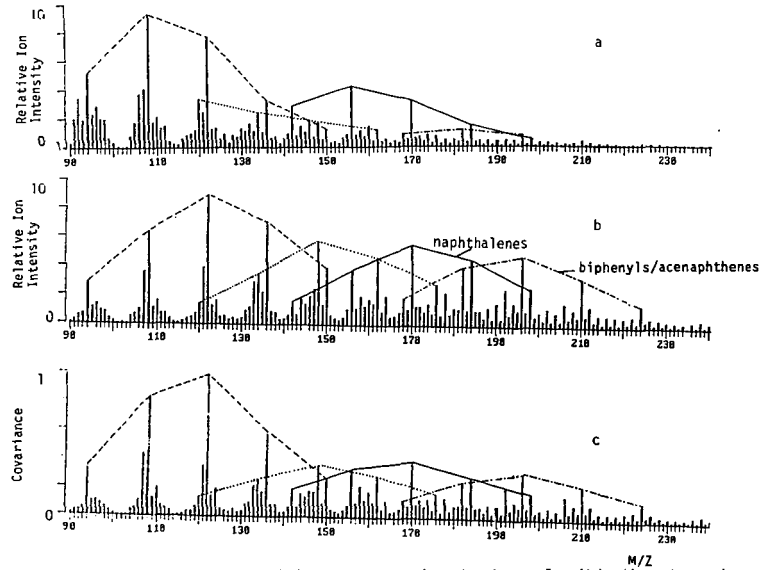

Figure 5. Pyrolysis mass spectrum of:(a) Upper Clarion hvAb coal; (b) direct probe mass spectrum of the corresponding tar, and (c) mathematically predicted spectrum.